Deutsches Institut für
Wirtschaftsforschung

DIW BERLIN

Discussion Papers

1036

Konstantin A. Kholodilin • Boriss Siliverstovs

Measuring Regional Inequality by Internet Car Price Advertisements: Evidence for Germany

Berlin, July 2010

IMPRESSUM

DIW Berlin
German Institute for Economic Research
Mohrenstr. 58
10117 Berlin

Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
http://www.diw.de

ISSN print edition 1433-0210
ISSN electronic edition 1619-4535

Papers can be downloaded free of charge from the DIW Berlin website:
http://www.diw.de/discussionpapers

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN:
http://ideas.repec.org/s/diw/diwwpp.html
http://www.ssrn.com/link/DIW-Berlin-German-Inst-Econ-Res.html

Measuring Regional Inequality by Internet Car Price Advertisements: Evidence for Germany*

Konstantin A. Kholodilin
DIW Berlin
Mohrenstraße 58
10117 Berlin
Germany
kkholodilin@diw.de

Boriss Siliverstovs
ETH Zurich
KOF Swiss Economic Institute
Weinbergstraße 35
8092 Zurich, Switzerland
boriss.siliverstovs@kof.ethz.ch

July 20, 2010

Abstract

We suggest to use Internet car sale price advertisements for measuring economic inequality between and within German regions. Our estimates of regional income levels and Gini indices based on advertisements are highly, positively correlated with the official figures. This implies that the observed car prices can serve as a reasonably good proxy for income levels. In contrast to the traditional measures, our data can be fast and inexpensively retrieved from the web, and more importantly allow to estimate Gini indices at the NUTS2 level — something that never has been done before. Our approach to measuring regional inequality is a useful alternative source of information that could complement the officially available measures.

Keywords: Car price advertisements; economic inequality; German NUTS1 and NUTS regions; Gini index; Internet
JEL code: C21; O47; R11

*This paper was first presented at a research seminar, which took place in the DIW Berlin on April 14, 2010. The paper further benefited from comments received at the BICEPS seminar in Riga and at the KOF Brown Bag seminar in Zurich. We are grateful to the participants of the seminars as well as to Vyacheslav Dombrovsky, Markus M. Grabka, d'Artis Kancs, and Christoph Moser for insightful discussions.

Contents

List of Tables

List of Figures

The automobile is not a luxury,
but a means of transport!

Ilya Ilf and Evgeniy Petrov
"The Golden Calf", 1930

1 Introduction

Internet offers an enormous amount of information that only now starts to be explored for practical purposes. Compared with the traditional information collection methods, the important advantage of using Internet data is that those can be retrieved much faster providing practically a real-time monitoring of the ongoing processes in the society. Thus, Internet can be viewed as an important alternative source of complimentary information to the traditional sources. In the pioneering study, Ginsberg et al. (2009)—that appeared online in November 2008—showed how one can use disaggregated Google searches filed by millions of users each day in order to study the intensity of influenza activity in the USA. Later several studies advocated the use of web queries for forecasting unemployment (e.g., Askitas and Zimmermann, 2009; Choi and Varian, 2009; D'Amuri and Marcucci, 2009) and private consumption (Kholodilin et al., 2010).

In this paper, we suggest to use information contained in car sale price advertisements placed on the web for measuring economic inequality in Germany both at the national level and within regions. Our paper is motivated by the following considerations.

First, the economic inequality of the households is an important characteristic of the welfare of a country or region. The societies experiencing too high inequality might be subject to more criminality, increased drug and alcohol consumption as well as political instability. Moreover, excessive inequality can have detrimental consequences for economic growth.[1] When, in addition, the inequality has a clear geographical pattern with pockets of poverty, on the one hand, and paradise islands, on the other hand, it can lead to the inter-regional tensions endangering the political unity of a country. The governments usually are trying at all cost to avoid such scenarios. Government policies to combat the inter-regional inequalities include fiscal federalism and structural policy.

However, the governments need certain indicators to measure their success in this respect. All these indicators are typically based on the survey data. The data are collected from a limited number of representative households, which are asked to fill the questionnaires including various questions concerning the expenditure and incomes of the households. While such a practice of data collection is widespread, there are a number of problematic issues: 1) Only a limited number of households are selected (invited) to participate, which makes the data not representative at the low regional level. For example, in Germany the household surveys are mainly conducted by the Socio-Economic Panel (SOEP), which typically collects the responses of approximately 11.000 households. Provided that there are 16 NUTS1 and 39 NUTS2 regions in Germany, that would imply that on average there are less than 700 and 200 observations per region, respectively. 2) The participation is voluntary and verification of supplied information is costly and may not always be pos-

[1] Thorbecke and Charumilind (2002) provide an extensive review of literature on inequality and its socio-economic impact.

sible. 3) The voluntary surveys suffer from the so-called "middle-class bias" (Becker and Hauser, 2003) when households with very high and very low income levels typically are not sufficiently represented. Keeping this in mind, we suggest an alternative source of information for measuring/monitoring regional inequality based on Internet car sale price advertisements.

Second, our choice of using cars for capturing economic inequality among regions is not purely incidental. A car is an affordable good that can be owned by everyone in the society from the poorest to the richest. More importantly, the cars are a specific durable good used not only for transportation but also for signalling the social status of an owner and his economic well-being. In this respect, cars are different from other durable goods like fridges or washing machines that are primarily used for what they are built for.

Third, in many respects Germany is an optimal example for our study. It is geographically and economically diverse country, where the government policy on lessening regional economic inequality have been on the top political agenda since the re-unification. Germans, as a car-building nation, love their cars and the well-maintained network of highways makes it pleasant to drive them. Moreover, cars is an affordable good in Germany. Indeed, in a country with 39.8 million of households there are about of 50.2 million registered cars (Federal Motor Transport Authority, in German — Kraftfahrt Bundesamt, KBA) and 81.2% of German households possess one or more cars, according to the SOEP 2008 survey[2]. In addition, it is also important for our approach that the Internet is also easily accessible to German households. According to the Eurostat, in 2009 about 71% of individuals in Germany used the Internet regularly (at least once a week). Moreover, this number is growing very fast: in 2006 this number was just 59%.

It is true that our approach is not exempt from criticisms. One can argue that by concentrating on the actual and potential car owners we exclude the poorest households from our sample. Indeed, the mean of the net calculated income of the households not having cars is about 1113 euros, whereas that of the households possessing cars is approximately 2521 euros. According to the ANOVA test, the difference between these two means is siginificant at 1% level[3]. Thus, the 18.8% of the households without cars are likely to be the poor households. However, even despite this weakness, our results are still reliable, as shown below.

The rest of the paper is structured as follows. In section 2, the data used in this study are described. Section 3 reports and discusses the results. Finally, section 4 concludes.

2 Data

The data have been downloaded during the period from 18th through 26th of May 2010 from a popular German website Mobile.de (http://www.mobile.de), where both new and used cars offered for sale are advertized. They include the following information: make, model, postal code, mileage, engine volume in liters and cubic centimeters, type of transmission (manual, automatic, etc.), year of the first registration, and offer price.

[2]Similar figures are reported in an independent study "Mobilität in Deutschland" (Mobility in Germany) conducted on behalf of the German Ministry for Transport, Construction, and Urban Development. According to this study, in 2008, 53% of German households possessed 1 car, 24% 2 cars, and 5% 3 or more cars. See Infas and DLR (2010), Figure 3.36, p. 60.

[3]All these computations are based on the data of the SOEP survey 2008

The postal code information was used to find the geographical coordinates (latitude and longitude) of each car's seller. Then, the price data were assigned to the respective NUTS1 and NUTS2 regions, given the information on their borders. The shapefile containing the geographical information on the regional borders was taken from the Eurostat.

In total, we collected about 671,000 unique car sale price observations offered in Germany, corresponding to about 7% of annual total car sales, given that in 2009 a total of 9.8 million cars was sold in Germany: 3.8 million new cars and 6.0 million cars sold at the secondary market.

3 Results

In this section, we conduct a descriptive analysis of the car price level both at the national and at the regional level. Selected descriptive statistics (average car prices and Gini indices) of the car prices for both the NUTS1 and NUTS2 regions are reported Tables 1 and 2. They characterize the intra- and inter-regional disparities.

Table 3 presents the estimated correlation coefficients between the data of the car prices advertisements, on the one hand, and official income estimates, on the other hand. The first line in the table reports the correlation coefficient for number of advertisements recorded in the respective NUTS1 or NUTS2 regions with the number of inhabitants. At both levels of disaggregation, the correlation is close to unity implying that the distribution of advertisements across German regions is proportional to the population of these regions and hence our data sample is geographically representative.

The middle panel of Table 3 reports the estimated correlation coefficient between different measures of income and the average car prices recorded for the respective NUTS1 or NUTS2 regions. The highest correlation coefficient, ρ, of 0.79 is reported for the national income (wage and property income) per capita and for the net primary income per capita at the NUTS1 level. At the NUTS2 level, the corresponding correlation coefficient between the average car price and the net primary income per capita is 0.71. The corresponding cross-plots as well as the maps depicting the geographical distribution of the average car prices are shown in Figures 1–2 and 3–4, respectively. The lower income levels of the East German Länder, or federal states (NUTS1 regions), seem to be well reflected in lower average car prices. This supports our assumption that car prices observed in a given region could serve as a good proxy for regional income levels, i.e., households living in a poor neighborhood tend to demand on average cheaper cars compared to the households that live in the more affluent regions. Thus, even given the equal unit prices for the same cars, the average car price for a richer region will be higher due to a different demand structure.

We also compared the values of the Gini index computed on the basis of the car prices to the official values based on the net equivalence income (Statistisches Bundesamt, 2010) available for the NUTS1 regions[4]. The estimated correlation coefficient is very high (0.88) reflecting very close correspondence between these two

[4]The official Gini indices are computed using the microcensus data. The microcensus is a population census, which is conducted every year at the reduced scale, covering about 1% of the total population of Germany, that is, around 800 thousand persons. This is, in fact, comparable to the size of our data sample. However, conducting the microcensus requires much money and time. Hence, our method can be considered as "quick and dirty" way of estimating the economic inequality compared to the microcensus. For more details on the microcensus see Statistisches Bundesamt (2008).

measures. The cross-plot and the geographical distribution of the Gini index for the NUTS1 regions are shown in Figures 5 and 6, respectively. The lowest Gini indices are observed in East German federal states pointing out to a lower inequality in the East, implying that even after 20 years since the re-unification the economic inequality in the East remains relatively low compared to that in the West. It should be noticed also that, apart from the Western federal states, the relatively high inequality is observed in the big cities: Berlin, Bremen, and Hamburg. Recall that during the cold war Berlin was divided in two parts, which apparently remain quite different from each other in terms of the welfare.

In Germany, several Gini indices are available for different measures of welfare summarized in Table 4. The smallest Gini indices (0.290-0.299) are computed for the net equivalence income of the households, whilst the largest Gini indices (0.683-0.703) are based on the wealth figures. Our estimates presented in the lowest row of the table are similar in magnitude to the Gini indices computed for the market equivalence income, for which the corresponding figures for NUTS1 are not available.

Our approach allows to estimate Gini index for the NUTS2 regions—something that never has been done before and therefore we cannot compare our results to the official ones. But, given our encouraging results for the NUTS1 regions, when these figures will be officially released it is very likely that they will resemble Figure 7.

4 Conclusion

We suggest an alternative indicator, which is based on the prices of the cars offered for sale in the Internet, for measuring economic inequality both at regional and national levels. Using Germany as an example we illustrate that our estimates of regional car price levels as well as of Gini indices have high, positive correlation with the official figures based on different measures of income. This implies that the observed car prices can serve as a reasonably good proxy for the income distribution. In contrast to the traditional measures, our data can be quickly and inexpensively retrieved from the Internet and, more importantly, allow to estimate Gini indices, or any other inequality measures, at the NUTS2 level—something that never has been done before.

An additional appealing feature of our approach is that it can provide a better indicator of economic inequality than the official estimates in countries with a relatively large share of informal economy. In those countries, the official estimates are likely to be severely downwards biased due to massive underreporting of earnings.

We conclude that our approach to measuring inequality appears to be a useful alternative source of information that could complement officially available measures but definitely more research is needed in order to verify our initial claim.

References

Askitas, N. and K. F. Zimmermann (2009). Google econometrics and unemployment forecasting. IZA Discussion Papers 4201, Institute for the Study of Labor (IZA).

Becker, I. and R. Hauser (2003). *Anatomie der Einkommenverteilung*. Berlin.

Choi, H. and H. Varian (2009). Predicting initial claims for unemployment benefits. Technical report, Google Inc.

D'Amuri, F. and J. Marcucci (2009). "Google it!" Forecasting the US unemployment rate with a Google job search index. MPRA paper, University Library of Munich, Germany.

Ginsberg, J., M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant (2009). Detecting influenza epidemics using search engine query data. *Nature 457*, 1012–1014.

Infas and DLR (2010). Mobilität in Deutschland 2008. Ergebnisbericht. Struktur – Aufkommen – Emissionen – Trends. Bundesministerium für Verkehr, Bau und Stadtentwicklung.

Kholodilin, K. A., M. Podstawski, and B. Siliverstovs (2010). Do Google searches help in nowcasting private consumption? A real-time evidence for the US. KOF Working papers 10-256, KOF Swiss Economic Institute, ETH Zurich.

Krause, P. and A. Schäfer (2005). Verteilung von Vermögen und Einkommen in Deutschland: Große Unterschiede nach Geschlecht und Alter. Wochenbericht nr. 11/2005, DIW Berlin.

Sachverständigenrat (2009). Jahresgutachten 2009/10.

Statistisches Bundesamt (2008). Mikrozensus 2006.

Statistisches Bundesamt (2010). Gini-Koeffizient der Äquivalenzeinkommen. http://www.amtliche-sozialberichterstattung.de/Tabellen/tabelleA3.html.

Thorbecke, E. and C. Charumilind (2002). Economic inequality and its socioeconomic impact. *World Development 30*(9), 1477–1495.

Appendix

Table 1: Inequality measures at the NUTS1 level

Bundesland	Short code	Number of ads	Population, thousands[1]	Average car price	Gini index	
					our[2]	official[3]
Baden-Württemberg	BW	90,939	10,748	14.1	0.495	0.278
Bayern	BY	111,907	12,497	14.6	0.491	0.286
Berlin	BE	21,496	3,432	10.9	0.502	0.296
Brandenburg	BB	17,916	2,516	11.0	0.443	0.256
Bremen	HB	4,418	660	13.1	0.501	0.294
Hamburg	HH	13,923	1,778	13.7	0.518	0.317
Hessen	HE	50,078	6,060	14.6	0.484	0.297
Mecklenburg-Vorpommern	MV	11,502	1,657	11.2	0.416	0.255
Niedersachsen	NI	69,160	7,945	12.5	0.456	0.279
Nordrhein-Westfalen	NW	140,261	17,893	14.1	0.479	0.289
Rheinland-Pfalz	RP	30,609	4,019	14.0	0.511	0.289
Saarland	SL	9,735	1,026	10.6	0.482	0.271
Sachsen	SN	36,714	4,177	12.1	0.448	0.246
Sachsen-Anhalt	ST	17,936	2,368	10.9	0.430	0.252
Schleswig-Holstein	SH	24,992	2,830	11.7	0.473	0.283
Thüringen	TH	19,723	2,257	11.4	0.428	0.243

[1] Number of inhabitants in 2007; source: Arbeitskreis VGR der Länder, http://www.statistik.baden-wuerttemberg.de/Arbeitskreis_VGR/ergebnisse.asp#BIP_K.
[2] Gini index computed for the car prices in May 2010.
[3] Gini index computed for the equivalence income (Statistisches Bundesamt, 2010).

Table 2: Inequality measures at the NUTS2 level

NUTS2 region	Short code	Number of ads	Population, thousands[1]	Average car price	Gini index of car prices
Stuttgart	DE11	35,139	4,007.3	13.7	0.484
Karlsruhe	DE12	22,525	2,736.3	14.4	0.471
Freiburg	DE13	18,115	2,195.4	15.8	0.540
Tübingen	DE14	13,954	1,807.3	12.0	0.487
Oberbayern	DE21	35,729	4,295.4	16.4	0.519
Niederbayern	DE22	12,923	1,193.9	13.7	0.455
Oberpfalz	DE23	10,991	1,087.3	13.2	0.447
Oberfranken	DE24	7,730	1,091.4	12.3	0.479
Mittelfranken	DE25	16,144	1,712.9	15.1	0.498
Unterfranken	DE26	10,743	1,336.2	13.3	0.475
Schwaben	DE27	17,032	1,787.6	13.6	0.472
Berlin	DE30	21,286	3,407.6	10.9	0.502
Brandenburg - Nordost	DE41	7,094	1,150.7	10.5	0.431
Brandenburg - Südwest	DE42	10,565	1,390.9	11.2	0.450
Bremen	DE50	4,209	663.3	13.1	0.501
Hamburg	DE60	13,659	1,761.7	13.7	0.518
Darmstadt	DE71	32,223	3,775.5	16.0	0.483
Gießen	DE72	8,179	1,055.1	11.5	0.460
Kassel	DE73	9,231	1,241.9	12.3	0.478
Mecklenburg-Vorpommern	DE80	11,291	1,686.7	11.2	0.416
Braunschweig	DE91	15,970	1,637.3	12.7	0.464
Hannover	DE92	16,623	2,159.4	11.9	0.463
Lüneburg	DE93	14,853	1,702.0	12.4	0.459
Weser-Ems	DE94	21,064	2,480.7	12.9	0.439
Düsseldorf	DEA1	39,163	5,212.8	14.7	0.483
Köln	DEA2	34,374	4,386.7	14.1	0.483
Münster	DEA3	21,470	2,616.8	12.7	0.453
Detmold	DEA4	16,955	2,062.5	13.5	0.487
Arnsberg	DEA5	27,502	3,733.1	14.4	0.476
Koblenz	DEB1	12,421	1,511.1	14.9	0.569
Trier	DEB2	5,664	515.7	15.1	0.415
Rheinhessen-Pfalz	DEB3	11,918	2,022.7	12.5	0.480
Saarland	DEC00	9,490	1,040.0	10.6	0.482
Chemnitz	DED1	15,783	1,583.4	12.9	0.430
Dresden	DED2	12,489	1,651.7	11.6	0.463
Leipzig	DED3	7,944	999.3	11.2	0.457
Sachsen-Anhalt	DEE0	17,701	2,427.6	10.9	0.430
Schleswig-Holstein	DEF0	24,692	2,835.3	11.7	0.473
Thüringen	DEG0	19,463	2,300.1	11.4	0.428

[1] Number of inhabitants in 2007; source: Arbeitskreis VGR der Länder, http://www.statistik.baden-wuerttemberg.de/Arbeitskreis_VGR/ergebnisse.asp#BIP_K.

Table 3: Estimated correlation coefficients

	NUTS1	NUTS2
Number of advertisements	0.995[1]	0.974
Gross regional product per capita	0.63[2][2009][3]	0.58 [2007]
Gross national income per capita	0.74 [2008]	-
Net national income (primary income) per capita	0.77 [2008]	-
National income (wage and property income) per capita	0.79 [2008]	-
Net balance of primary income, per capita	0.79 [2008]	0.71 [2007]
Disposable income, net (uses)	0.73 [2008]	0.66 [2007]
Gross wage per employee	0.73 [2009]	0.60 [2007]
Private consumption per capita	0.70 [2008]	-
Gini index (net equivalence income)	0.88[4]	-

[1] Correlation between the number of advertisements and number of inhabitants.
[2] Correlation between the average car prices and income measures.
[3] In squared brackets the reporting year is indicated.
[4] Correlation between the Gini indices: car-price-based and based on net equivalence income (Statistisches Bundesamt, 2010).
[5] Sources: all income measures are taken from Arbeitskreis VGR der Länder, http://www.statistik.baden-wuerttemberg.de/Arbeitskreis_VGR/ergebnisse.asp#BIP_K; own calculations.

Table 4: Gini indices for Germany

Welfare measure	year	Gini index
Market equivalence income[1]	2007	0.473
Net equivalence income[1]	2007	0.290
Net equivalence income[2]	2008	0.290
Net equivalence income[3]	2002	0.299
Gross wealth[4]	2002	0.703
Net wealth[4]	2002	0.683
Car-price-based[5]	2010	0.481

[1] Sachverständigenrat (2009), household income per person, in 2005 prices, with imputed rent.
[2] Statistisches Bundesamt (2010).
[3] Krause and Schäfer (2005), household income per person, in 2002 prices, without imputed rent.
[4] Krause and Schäfer (2005).
[5] Own calculations.

Figure 1: Average car prices vs per capita net balance of primary income, NUTS1; estimated correlation coefficient $\rho = 0.79$

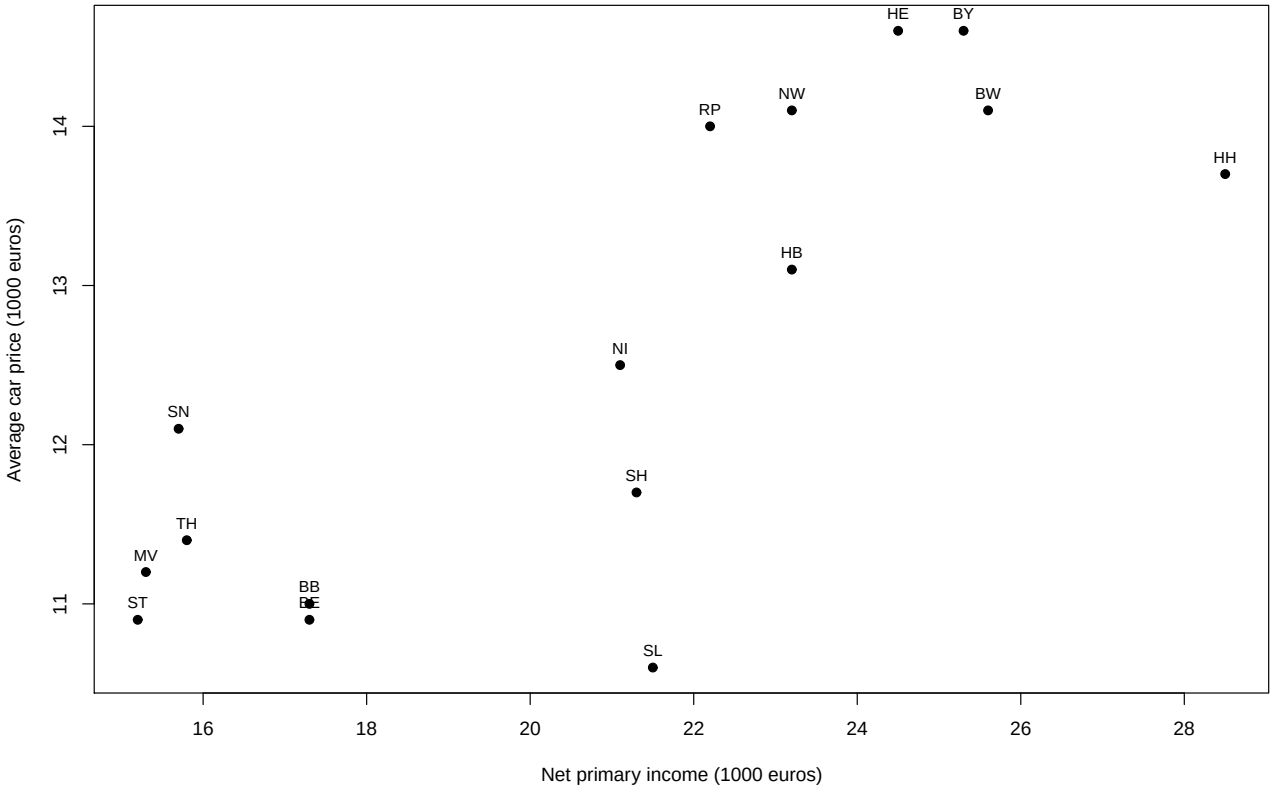

Figure 2: Geographical distribution of average car prices (1000 Euro); NUTS1

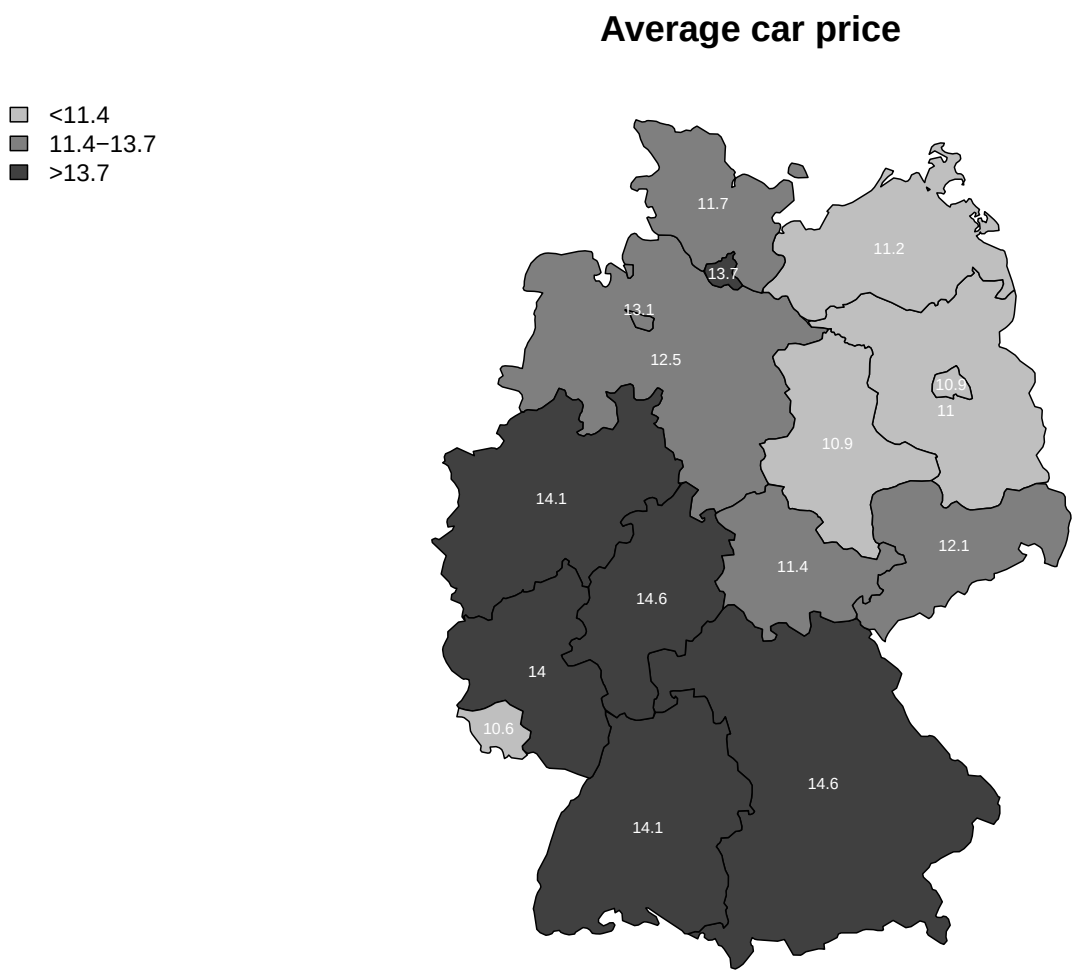

Figure 3: Average car prices vs. per capita net balance of primary income, NUTS2; estimated correlation coefficient $\rho = 0.71$

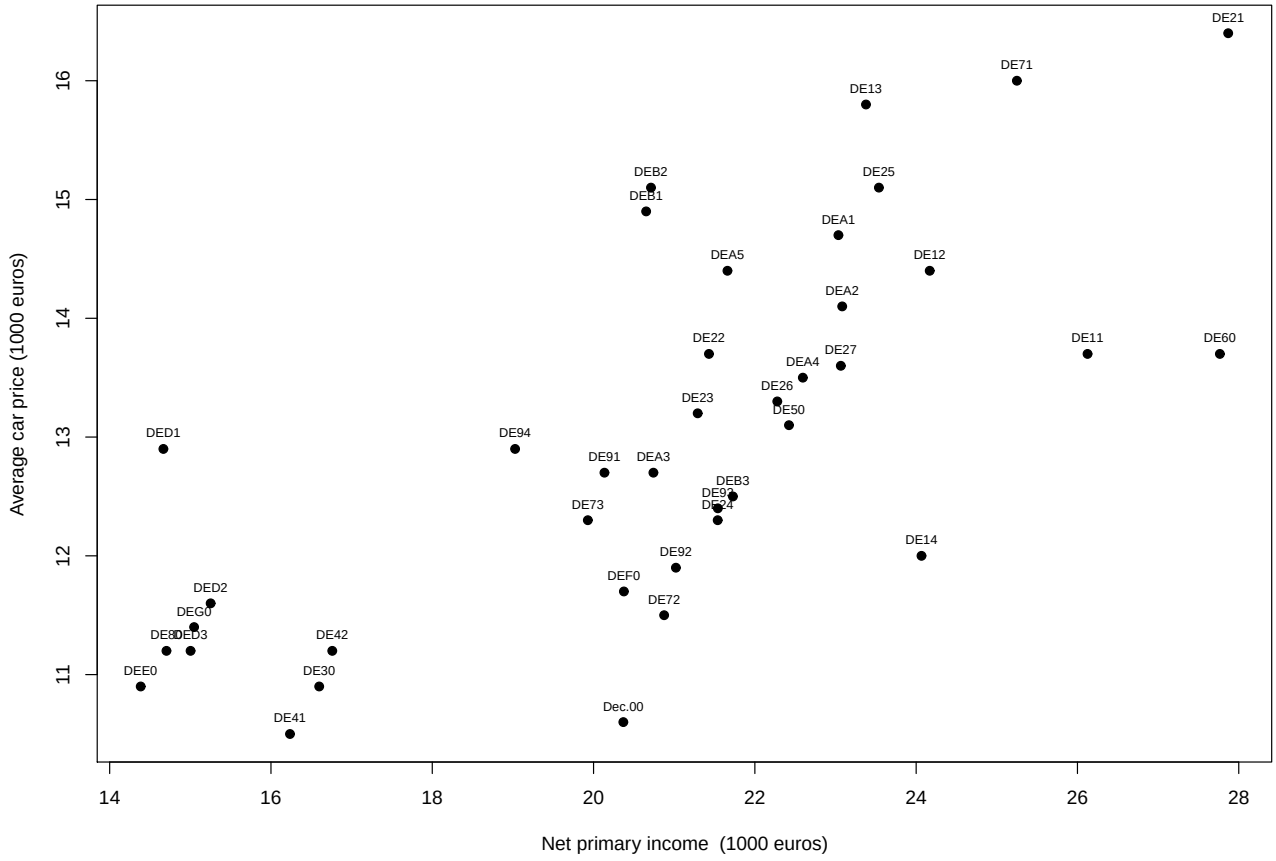

Figure 4: Geographical distribution of average car prices (1000 Euro); NUTS2

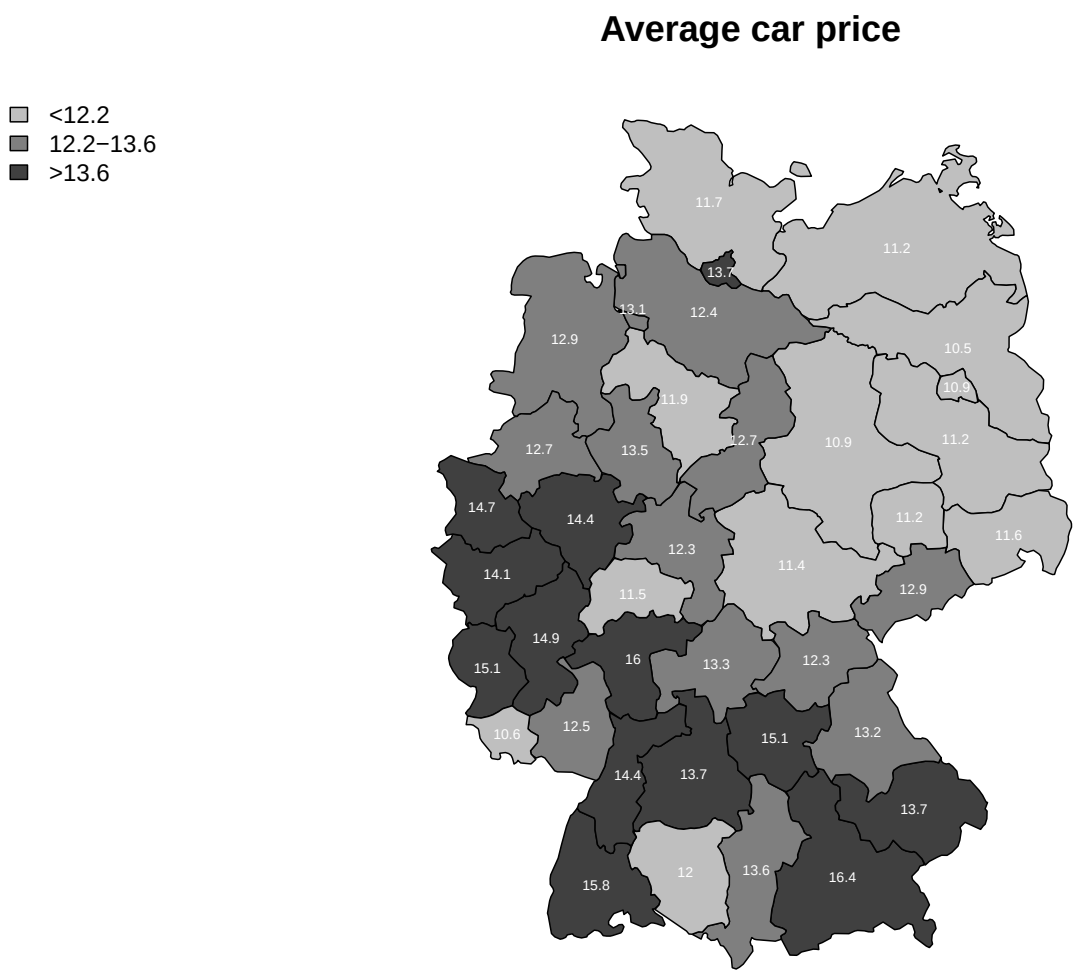

Figure 5: Gini coefficient (car-price-based vs. official estimates), NUTS1; estimated correlation coefficient $\rho = 0.88$

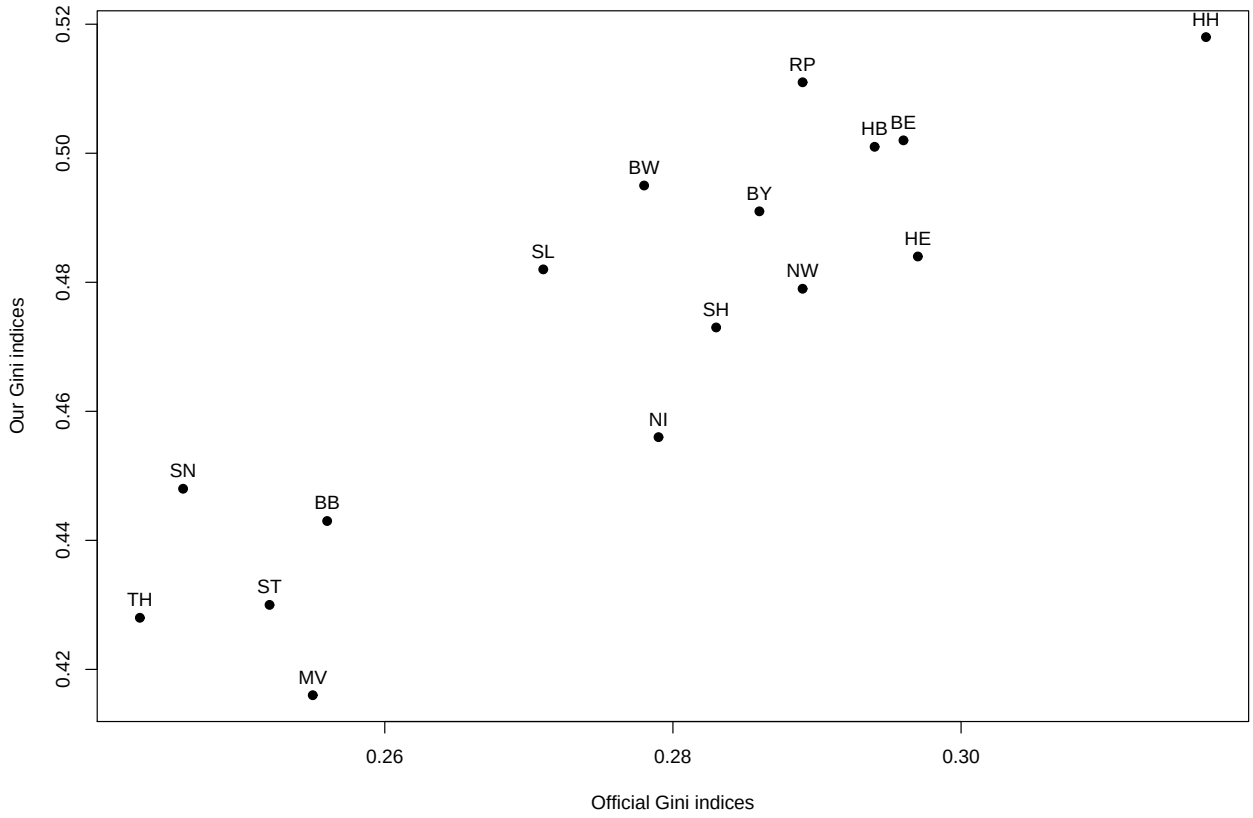

Figure 6: Geographical distribution of Gini coefficient (car-price-based); NUTS1

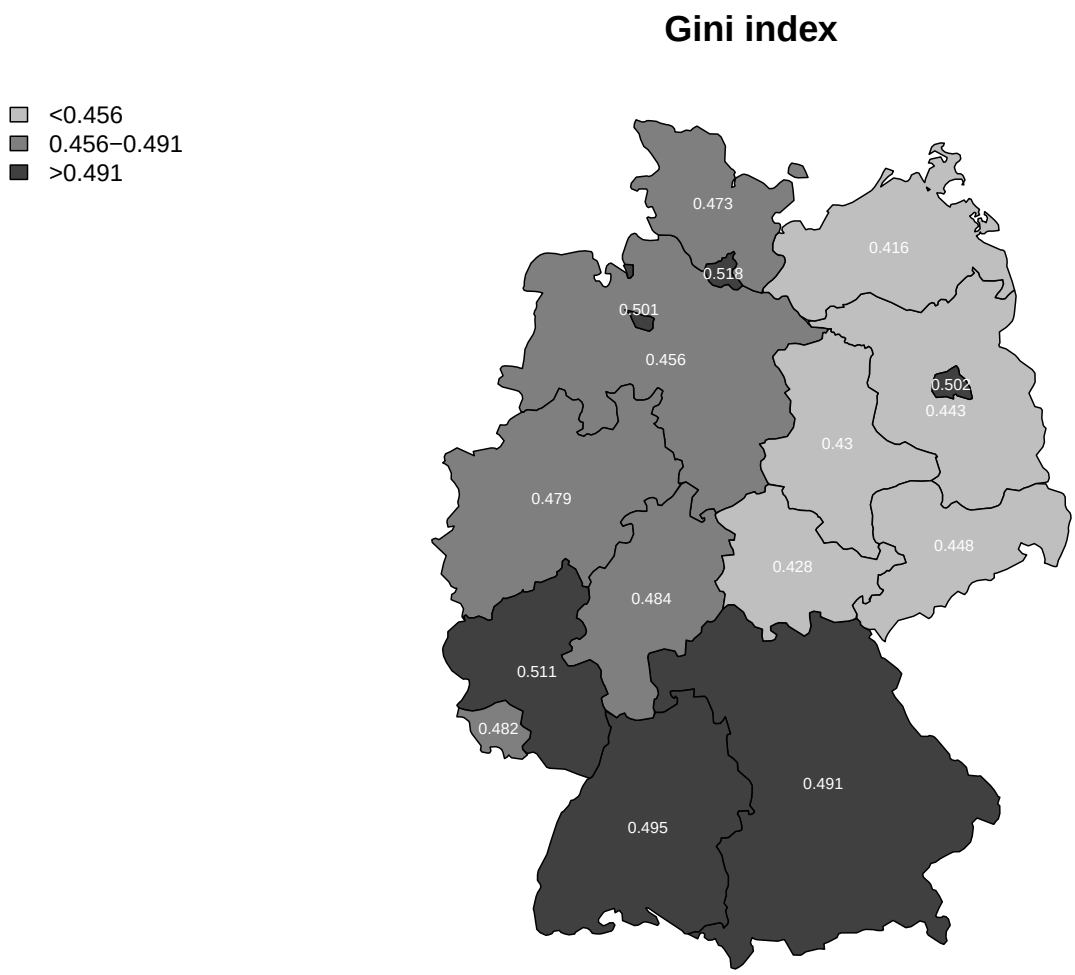

Figure 7: Geographical distribution of Gini coefficient (car-price-based); NUTS2